

UM DEBATE SOBRE A NATUREZA DOS PRIMITIVOS DE PERCEPÇÃO DA FALA

On the controversy of speech perception basis

Gustavo Nishida*

RESUMO

O objetivo deste trabalho é apresentar um “debate” ocorrido entre John Ohala e Carol Fowler em 1996, no qual discutiram qual é a natureza da percepção da fala. Sua apresentação é importante, pois há um hiato com relação às discussões sobre a natureza da percepção de fala. O problema é colocado na década de 50 (com os achados de Liberman e colegas nos Laboratórios Haskins) e somente é retomado em 1996 com esse “debate”.

A análise dessa troca de textos mostra que os primitivos perceptuais seguem as propostas de primitivos de análise das teorias fonológicas, uma vez que no Estruturalismo acreditava-se numa concepção auditiva (acústica) sobre a percepção, enquanto ela passa a ser de natureza abstrata e mental no Gerativismo e real e ecológica no arcabouço dinâmico.

É sobre esse panorama histórico que o debate entre Ohala e Fowler ilustra de maneira pontual que não há consenso entre os pesquisadores sobre a natureza da percepção da fala. A nossa proposta é a de que esse fato existe porque cada um dos debatedores se situa em um recorte epistemológico que, por sua vez, considera fenômenos distintos como relevantes para sustentar cada uma de suas propostas teóricas, a saber: enquanto Ohala sustenta seus argumentos com base em evidências fonológicas de natureza acústica, Fowler salienta os aspectos articulatórios presentes na percepção e produção da fala.

Por fim, ao avaliar as premissas da Teoria Motora da Percepção da Fala (LIBERMAN; MATTINGLY, 1985), tentamos “esboçar”

* Universidade Tecnológica Federal do Paraná.

quais seriam as suas respostas ao debate. Seguindo as bases da Fonologia Gerativa Padrão (CHOMSKY; HALLE, 1968) e da Modularidade da Mente (FODOR, 1983), seus questionamentos sugerem pontos distintos de discussão e de interesse. Trata-se, então, em um exemplo hipotético sobre como as teorias constroem seus objetos de maneira distinta.

Palavras-chave: teorias de percepção da fala; teoria motora de percepção da fala, teoria do realismo direto de percepção da fala.

ABSTRACT

The goal of this paper is to present a discussion occurred in 1996 between John Ohala and Carol Fowler. The theme of that issue was the ontology of speech perception. The presentation of that discussion is important because there aren't enough studies about this topic. Basically, the problem was brought to light in 1950 with the Haskins Laboratories research and it was reconsidered in 1996 by these researchers.

We decided to make an analysis of Ohala's and Fowler's papers to show that the perceptual primitives choice follows the phonological theories proposals. In that debate, we can observe the fact that in Structuralism speech perception was based on auditory and acoustic features. On the other hand, speech perception is considered abstract and based on mentalist forms in the Generative approach and realistic and ecological in the dynamic approach.

The debate shows that there is no consensus about the ontology of speech perception. We propose that happens because each point of view is supported by a distinct epistemological basis. For example, while Ohala supported his arguments on acoustics phenomena, Fowler used only articulatory data.

Ultimately, we decided to consider the premises of the Motor Theory of Speech Perception (LIBERMAN; MATTINGLY, 1985) and his basis (the Sound Pattern of English – CHOMSKY; HALLE, 1968 – and The Modularity of Mind – FODOR, 1983) to make a “hypothetical pronunciation”, wondering what the Motor Theory would answer. The answers shows that each theory considers the speech perception differently.

Keywords: speech perception theories; motor theory of speech perception; direct realist theory of speech perception.

1. INTRODUÇÃO

Em trabalho anterior (NISHIDA, 2012), dedicamo-nos a apresentar as diferentes maneiras de conceber a natureza dos primitivos de percepção da fala. De modo geral, os trabalhos que tentaram responder a essa questão a consideraram ou como um primitivo de análise composto por propriedades estritamente acústicas (e.g. PISONI, 1973, PISONI; TASH, 1974 e COLE; SCOTT, 1974) ou como de base articulatória (e.g. LIBERMAN; MATTINGLY, 1985 e FOWLER, 1984).

Nesse mesmo trabalho, tentamos mostrar que a adoção de cada um desses primitivos perceptuais está vinculada à proposição de novas teorias fonológicas que, a seu tempo, sugerem novos primitivos para o tratamento do nível fonológico das línguas. Vale ainda salientar que tais teorias fonológicas, como a Fonologia Gerativa (CHOMSKY; HALLE, 1968) e a Fonologia Articulatória (BROWMAN; GOLDSTEIN, 1986, 1989, 1990, 1992), estão relacionadas a Programas de Investigação Científica distintos (cf. LAKATOS, 1979), os quais promovem olhares diferentes sobre quais são os objetos de interesse para cada uma das teorias que tratam dos aspectos fônicos da linguagem.

Por exemplo, no período Estruturalista, há uma busca pela invariância supostamente presente no sinal acústico, uma vez que o sinal de fala teria a função de transmitir informações de um emissor a um destinatário, as quais atuam em um dado sistema de comunicação. É desse período a proposição de que os fonemas eram compostos por feixes de traços de base acústica (JAKOBSON; FANT; HALLE, 1952) que promoviam distinção fonêmica no sistema fonológico das línguas. Com isso, esperava-se que tais propriedades acústicas estivessem presentes na realização dos fones e que elas fizessem com que os aspectos distintivos de um dado fonema fossem perceptualmente audíveis (cf. NISHIDA, 2010).

Diante, então, do fracasso da busca pela invariância acústica (tal como apresentada por LIBERMAN, 1957), os pesquisadores que se debruçavam sobre os estudos de percepção da fala começaram a estipular que a sua natureza era de base articulatória e não acústica. Ou seja, se o domínio acústico não apresenta propriedades invariantes para compor os fonemas, o domínio articulatório deveria ser invariante. Caso contrário, seria impossível a transmissão de informação em um dado sistema comunicativo.

Desde a proposição da *Early Motor Theory* (LIBERMAN, 1957), os trabalhos de Liberman e colaboradores nos Laboratórios *Haskins* precisavam de um arcabouço teórico que possibilitasse o desenvolvimento da proposta de que a percepção era guiada pela articulação. Tais marcos teóricos (a Fonologia Gerativa Padrão de CHOMSKY; HALLE, 1968 e a Modularidade da

Mente de FODOR, 1983) só surgiram anos após esses primeiros achados, e isso possibilitou a formulação da Teoria Motora da Percepção da Fala (LIBERMAN; MATTINGLY, 1985), doravante TM. Nessa teoria, os primitivos perceptuais são de base articulatória abstrata, i.e., trata-se de primitivos que representam os comandos motores pretendidos pelos falantes que são recuperados através de um módulo fonético que é alimentado pelos sons produzidos e veiculados no ar durante a fonação.

Uma outra teoria de percepção da fala (a Teoria do Realismo Direto de Percepção da Fala – FOWLER, 1984 e seguintes – doravante TRD) começa a ser ensaiada a partir dos avanços promovidos nas técnicas de análise acústica e articulatória da fala. Tais avanços possibilitam que se tente buscar como as características coarticulatórias presentes na produção dos sons influencia a percepção de vogais e consoantes. Nessa perspectiva, assume-se a coarticulação como uma característica inerente à fala e também que ela estaria envolvida na sua percepção. Trata-se de uma abordagem bem distinta das anteriores, pois toma a fala dinamicamente, i.e., a fala passa a ser considerada no seu desenvolvimento temporal.

A incorporação do tempo nas análises sobre a fala possibilita a adoção de um novo primitivo de análise fonológica (o gesto articulatório – BROWMAN; GOLDSTEIN, 1992), que, por sua vez, fomenta o olhar para a percepção de uma maneira não puramente simbólica (tal como na TM), mas sim realista (com base em movimentos que de fato acontecem no momento da produção da fala) e dentro de um sistema de desenvolvimento ecológico da espécie humana (GIBSON, 1966).

Como se pode notar, em cada um desses Programas de Investigação Científica (a saber, Estruturalista, Gerativista e Dinamicista), há a proposição de teorias de Fonologia e de produção e percepção da fala que sustentam seus respectivos núcleos epistemológicos.

Em outro trabalho (NISHIDA, 2012), descrevemos detalhadamente a construção de cada uma das abordagens teóricas, de modo que foi possível vincular cada uma delas a um programa de investigação científica em linguística, para que não houvesse dúvidas com relação aos domínios epistemológicos de cada proposta.

No entanto, as teorias de percepção da fala se encontram em concorrência, uma vez que não há unanimidade com relação a qual seria a natureza dos primitivos perceptuais. Isso se verifica com a polêmica na qual Fowler (1996) e Ohala (1996) se envolvem. Esse é motivo pelo qual apresentaremos essa querela neste texto. Nas subseções que se seguem, indicaremos os aspectos teóricos que impossibilitam o fechamento da polêmica de modo a sustentar a incompatibilidade epistemológica entre essas teorias.

Primeiro, passaremos à apresentação da proposta e dos argumentos de cada um dos envolvidos na polêmica. Em seguida, realizaremos um exercício de responder ao debate a partir das premissas da TM, para ilustrar que cada uma das teorias considera a percepção da fala de uma maneira distinta. Por fim, sumaremos os motivos pelos quais a polêmica não fecha.

2. UMA POLÊMICA SOBRE A NATUREZA DOS PRIMITIVOS DE PERCEPÇÃO DA FALA

Nesta seção, apresentaremos uma polêmica em torno da natureza da percepção da fala. Trata-se do posicionamento de base auditiva (ou de base acústica) defendido por Ohala (1996) contra as propostas de base articulatória de Fowler (1996). Após apresentar a polêmica, realizaremos uma análise das propostas salientando a diferença que há entre cada uma das teorias de percepção da fala; mostraremos também o papel que o acústico assume em cada uma delas. O ponto de chegada da nossa proposta é mostrar que a polêmica não se desenvolve justamente pelo fato de os programas de pesquisa defenderem seus pressupostos teóricos a partir de um local específico bem demarcado.

A escolha dessa polêmica em nossa análise não foi tomada ao acaso. Primeiramente, é preciso ressaltar que as discussões acerca da natureza da percepção da fala têm seu início na década de 1950 com os primeiros achados de Liberman e colaboradores nos Laboratórios Haskins. No entanto, as discussões não eram “travadas” em forma de “troca de textos”. Eram, antes de tudo, discussões acerca de achados experimentais que colocavam em xeque hipóteses (ou concepções) ainda insipientes sobre a percepção da fala. Basta nos lembrarmos dos relatos e das discussões sobre o fato de que os “invariantes” acústicos da fala eram determinados na base da “tentativa-e-erro” com o auxílio de um aparelho que convertia espectrogramas desenhados à mão em som: o *Pattern Playback*. Até a proposição da TM, as discussões sobre a natureza da percepção da fala eram realizadas sem que houvesse um “oponente” bem determinado. Só por esse aspecto, a troca de textos entre Ohala e Fowler já merece nossa atenção.

Outro motivo pelo qual tomamos esse “debate” como objeto de estudo é o fato de se tratar de um discurso polêmico sobre a natureza da percepção da fala, i.e., trata-se do ambiente ideal para opor ideias e disputar “território” para uma determinada teoria. Entretanto, o ponto interessante dessa polêmica é justamente o fato de ela ser colocada e não se desenvolver. A nossa hipótese é a de que a discussão não ganha atenção pelo fato de as teorias estarem assentadas sobre terrenos epistemológicos distintos, o que faz com que a polêmica não chegue a lugar algum¹.

1 Há pelo menos duas outras maneiras de considerar o quadro apresentado. Uma

Passemos, então, à polêmica e depois à análise e discussão sobre o motivo pelo qual a polêmica não fecha.

2.1 ENTRE O ACÚSTICO E O ARTICULATÓRIO: OHALA VERSUS FOWLER

Ohala (1996) argumenta em seu texto que “a percepção da fala não envolve a recuperação das articulações que geram o sinal acústico” (Ohala, 1996, p.1718). Para o autor, recuperar a articulação subjacente à fala não é uma tarefa essencial para entender a sua percepção. Em princípio, ela é importante apenas para o trabalho do cientista da fala. Tendo isso em vista, ele vai sugerir que as teorias de base motora – tais como a Teoria Motora de Percepção da Fala (LIBERMAN; MATTINGLY, 1985) e a Teoria do Realismo Direto da Percepção da Fala (FOWLER, 1986, 1996) – são uma maneira extravagante de encarar a percepção da fala.

Seus argumentos em favor de um olhar auditivo sobre a percepção levam em conta: 1) algumas evidências fonológicas; 2) mudanças sonoras; e 3) algumas reflexões que indicam a falta de plausibilidade dessas teorias. Vamos a eles.

Ohala sugere que, fonologicamente, há características nos sons das línguas que são capturadas mais satisfatoriamente por traços de base acústica² tais como [grave] e [flat] na proposta dos *Preliminaries of Speech Analysis* (JAKOBSON; FANT; HALLE, 1952). Sons [+graves] apresentam predominantemente energia de produção localizada na porção mais baixa e inicial do espectro. Por exemplo, sons labiais e velares seriam identificados como [+grave]; em oposição, apicais e palatais não. Tais dados mostram que há “um conjunto descontínuo³ de um ponto de vista articulatório” (OHALA,

é considerar que a discussão sobre o tema não se desenvolve do mesmo modo que a polêmica. De certa forma, isso explicaria a nossa sensação de que a polêmica não se desenvolve, mas as teorias, sim. Um outro aspecto seria o fato de as discussões sobre as fronteiras entre as disciplinas Fonética e Fonologia serem ainda muito recentes. Podemos dizer que, salvo alguns espaços simpatizantes (cf. ALBANO, 2001; SILVA, 2002), a Fonética ainda permanece na periferia do núcleo duro das questões linguísticas.

2 Poderíamos nos perguntar se os traços [grave] ([bemol]) e [flat] são de base acústica ou auditiva. Esse ponto não fica muito claro para a proposta dos *Preliminaries*, uma vez que o tratamento instrumental para a determinação dos traços a partir das características visuais (via espectrografia) do sinal acústico é realizado para validar o que se ouve como distintivo. Talvez seria melhor considerá-los como acústico-auditivo, já que a percepção é a audição das características acústicas que promovem distinções fonológicas.

3 Ohala considera o traço [+grave] como articulatoriamente descontínuo porque ele especifica regiões articulatórias distintas (e opostas) do trato vocal. Cabe-nos perguntar qual foi o motivo que levou Ohala a iniciar a sua análise pelos traços positivos, uma vez que eles poderiam categorizar a porção medial do trato, que caracterizaria um contínuo articulatório. A mesma reflexão cabe para o traço [flat]. Ohala não discute tais motivos e tampouco desenvolve as questões subjacentes referentes à classe de som. Contudo, como consideramos que se trata de um problema para todas as teorias fonológicas, decidimos não entrar nessa discussão aqui.

1996, p.1721). O mesmo traço consegue explicar porque algumas vogais do inglês não sofreram alteração no *English Vowel Shift*. Por exemplo, historicamente o ditongo [aw] do inglês veio de um [u:] (e.g., *mouse* [maws] < [mu:s] e *south* [sawθ] < [su:θ]), mas hoje o ditongo [aw] apenas ocorre antes de sons que têm o traço [-grave] (e.g., *shout*, *loud*, *house*, *howl*, *town* – [t], [d], [z], [l] e [n], respectivamente) e nunca antes de [+grave] (e.g., *[awp], *[awm], *[awk]). A razão pela qual isso ocorre é devido à possibilidade de consoantes [-grave] produzirem um traçado descendente do F2 com relação à vogal [u]. Em contrapartida, consoantes [+grave] não produzem essa “queda” de F2 pelo fato de possuírem um F2 similar ao de [u], não conseguindo criar a vogal ditongada. Ou seja, Ohala argumenta em favor de uma abordagem acústica, pois a percepção de um ditongo está associada ao traçado de F2 e não às articulações envolvidas na sua produção.

Na mesma esteira de descontinuidade, do ponto de vista articulatorio está o traço [flat]. Ele se aplica aos sons retroflexos e aos que apresentam articulação secundária nos lábios, véu (e úvula) ou faringe. Trata-se também de um conjunto de articulações descontínuas, pois não se incluem apicalizações ou palatalizações.

Ainda dentro da gama de evidências fonológicas apontadas por Ohala, há o argumento de que, como os sons são veiculados na fala, espera-se que as diferenciações sejam encontradas no domínio acústico. Para endossar essa hipótese, o autor aponta para os achados de Stevens (1989) e Ohala (1983):

Parece que os inventários dos sons da fala exploram uma relação não linear entre a produção da articulação e dos sons para fazer com que os elementos dos sons da fala se diferenciem. (OHALA, 1996, p. 1719⁴)⁵

Os argumentos sobre a mudança sonora nas línguas do mundo caminham na mesma direção da ideia de que elas ocorrem por haver alguma variação no aspecto acústico-auditivo e não na articulação. Em seu texto, há a apresentação de dados sobre: 1) as labiais palatalizadas (do tcheco padrão, italiano de Roma e proto-bantu), que são confundidas com alveolares (realizadas, respectivamente, como forma inovadora no boêmio do leste, italiano genovês e zulu) – cf. OHALA, 1978 *apud* OHALA, 1996, p.

4 “[The answer is that] speech sound inventories exploit the non linear relation between articulation and the sound produced in order to make the elements of speech *sound* different” (Ohala, 1996, p.1719).

5 A tradução dos originais é nossa.

1722; 2) a nasalização espontânea do sânscrito para o hindi e do francês para o bretão (OHALA; OHALA, 1993 *apud* OHALA, 1996, p.1722); e 3) as nasais epentéticas antes de oclusivas sonoras na mudança do hindi antigo para o moderno.

Por fim, há os argumentos contra a falta de plausibilidade das teorias que propõem que a percepção da fala é guiada pela articulação. Ohala argumenta que há estudos com animais e com crianças em período anterior à fala que comprovam a sua capacidade de categorizar os sons da fala. Para ele, é muito difícil que nesses casos esteja ocorrendo a recuperação das articulações utilizadas. Para além disso, Ohala se pergunta se conseguimos recuperar todas as articulações utilizadas quando ouvimos uma corneta tocando, uma vez que seu som é produzido a partir da variação de pressão da respiração e da tensão dos lábios⁶.

Um outro ponto interessante apontado pelo autor é o fato de que aprendizes de segunda língua e pessoas em terapia de fala aprendem a produzir os sons de uma língua a partir de explicações articulatórias. No entanto, tais tarefas tomam muito tempo e esforço, sugerindo que a aprendizagem de sons talvez não se dê através do conhecimento articulatório e do conhecimento sim acústico-auditivo.

Finalmente, há um último comentário do autor, denominado *Chicken Little Inquiry*:

Chicken Little (ou em algumas versões, Henny Penny) é o personagem principal de um desenho infantil familiar à maioria dos falantes de inglês. Um dia, Chicken Little recebeu um grande sopro na cabeça e então colocou todo o celeiro em pânico com a sua teoria de que o céu estava caindo. A resolução da estória não envolve experimentos que testam se o céu, de fato, estava caindo; envolveu uma investigação sobre a razão pela qual Chicken Little pensava que o céu estivesse caindo. A evidência imediata foi um inchaço no topo de sua cabeça. Onde estava ele quando o dano ocorreu? Embaixo de um carvalho. E eis que uma grande bolota foi encontrada no mesmo local! A moral da história - que não ficou tão explícita assim - é de que, antes de investir muito tempo e esforços em experimentos caros para avaliar uma hipótese qualquer, devemos primeiro focalizar a motivação para a hipótese construída. Então devemos nos perguntar se as observações que originaram a hipótese poderiam ser tratadas por outras hipóteses menos extravagantes (OHALA, 1996, p.1724)⁷

6 Talvez consigamos sim recuperar essa variação de pressão da respiração e pressão dos lábios quando ouvimos uma pessoa tocando uma corneta. Basta pensarmos que, de modo geral, conseguimos imitar os sons de um instrumento dessa natureza. A destreza na afinação e fidelidade de timbre depende do imitador da mesma forma que o som do instrumento depende do músico que está tocando.

7 "Chicken Little is the principal figure in a children's story familiar to most English

Com essa anedota, o autor conclui que as teorias de base motora são em realidade uma hipótese extravagante sobre o fato de haver falta de invariância entre as unidades linguísticas assumidas na fala e sua manifestação acústica. Para Ohala, esse argumento é do tipo “Como não encontro isso aqui, ele deve estar ali!” (Ohala, 1996, p.1724⁸).

Mais além (com base em STEVENS; BLUMSTEIN, 1978), o autor sugere que “os invariantes estão presentes no sinal acústico, as pessoas apenas estão olhando para os traços errados” (OHALA, 1996, p.1724⁹) e que “algumas relações de ordem alta entre certos parâmetros acústicos mostrarão uma relação invariável com as unidades linguísticas” (SUSSMAN *et al.*, 1991 *apud* OHALA, 1996, p.1724¹⁰). Para ele, não há um único traço invariante para cada unidade. Ao contrário, há uma coleção de traços; a identidade de uma dada unidade é determinada se algum número crítico desses traços estiver presente. Ele encerra comentando que as teorias motoras estão considerando os fonemas como unidades de codificação, mas difones poderiam ser candidatos melhores.

A resposta de Fowler (1996) a Ohala tem três objetivos: 1) mostrar que há diferenças entre a Teoria Motora (TM) e a Teoria do Realismo Direto (TRD), pois ele as considerou como equivalentes; 2) apresentar argumentos a favor da percepção dos gestos articulatórios; e, por fim, 3) trazer argumentos que comprovem que os pontos levantados por Ohala não refutam a teoria. Com isso, passaremos à apresentação dos pontos principais de cada objetivo.

Basicamente, o desenvolvimento da TRD está ligado com o desenvolvimento das teorias de produção da fala e da fonologia (como a Fonologia Articulatória – BROWMAN; GOLDSTEIN, 1992) e com uma teoria mais geral de percepção direta (GIBSON, 1966, 1979). Para os estudiosos deste viés teórico, a percepção dos gestos não é logicamente necessária para ouvintes/falantes habilidosos – ela é biologicamente necessária. Isto é, a percepção da fala somente é possível se ela for direta (sem intermediários transformacionais) e se consideramos que há um sistema perceptual com funções universais moldados pela seleção natural.

speakers. Chicken Little received a tremendous blow on the head one day and then set the entire barnyard into a panic with her theory that the sky was falling. The resolution of the story did not involve experiments testing whether the sky was, in fact, falling; it involved an inquiry as to why Chicken Little thought it was falling. The immediate evidence was a swelling on the top of her head. Where was she when the injury occurred? Under an oak tree. A large acorn was found on the very spot! The moral of the story—not stated quite so explicitly—is that before investing a lot of time and effort in costly experiments evaluating a given hypothesis, we should first look at the motivation for the hypothesis having been made. Then we should ask whether the observations prompting the hypothesis might be accounted for by other, less extravagant, hypotheses”. (Ohala, 1996, p. 1724).

8 “[The reasoning here is of the sort] ‘I can’t find it here, so it must be over there’” (Ohala, 1996, p.1724).

9 “[...] the invariants are present in the acoustic signal, one was just looking at the wrong features (Stevens and Blumstein, 1978).” (Ohala, 1996, p.1724)

10 “[...] some higher-order relations among certain acoustic parameters will show an invariant relation to linguistic units (Sussman et al., 1991).” (Ohala, 1996, p.1724)

Essa última é biologicamente fundamental porque é por esse sistema perceptual que os animais conseguem reconhecer seu meio; conforme a proposta, isso foi moldado pela evolução. Ainda mais, os sistemas utilizam estruturas em mídia (como o ar e a água), causadas por eventos no ambiente para promover a sua informação, i.e., luz para a visão, pele para o tato e ar para a audição. Os seres, então, perceberiam os componentes de seu meio que causaram a estrutura.

Fowler traz alguns achados experimentais como evidências para a sua proposta. Um deles é o chamado efeito *McGurk* e vamos nos deter, por ora, apenas nele. Nesse experimento, os pesquisadores (MCGURK; MACDONALD, 1976) sincronizaram a imagem da boca de um falante produzindo a sílaba [da] com o sinal acústico de um [ma]. A maioria das pessoas (92% dos informantes) respondeu que a imagem havia produzido um [na], i.e., os informantes perceberam o ponto de articulação visualmente e o modo de articulação auditivamente¹¹. Os autores sugeriram que as pessoas percebem gestos (de maneira geral) e que alguns gestos são percebidos visualmente e outros auditivamente. Outro trabalho (MASSARO, 1987) sugeriu que em realidade os falantes associam as experiências (armazenadas na memória) de ver e ouvir as pessoas falando com uma imagem mental de um fonema (ou sílaba). Isso, para ele, explicaria o fato de as pessoas conseguirem ler lábios, por exemplo. No entanto, em outro experimento (FOWLER; DEKLE, 1991), foi possível mostrar que tais associações não ocorrem.

Para tanto, eles partem do seguinte princípio: se o efeito *McGurk* emerge da associação, na memória, entre um sinal (ou gesto) e um som produzido por um falante, então um efeito análogo pode ocorrer quando há integração entre o tato (um sistema de base motora e articulatória) e a fala. Ou seja, os experimentadores queriam testar se o efeito *McGurk* ocorre quando um sujeito (de olhos fechados) ouve a fala e toca a face de alguém. Para testar essa hipótese, eles realizaram dois experimentos.

No primeiro, sintetizaram um contínuo de sílabas de /ba/ até /ga/. A tarefa consistia em identificar a sílaba que eles ouviam e, então, identificar de alguma maneira a sílaba que eles viam, pois à medida que era apresentado um estímulo auditivo de /ba/ ou /ga/ era apresentada na tela do computador as sílabas BA ou GA. Nesse experimento, foi possível notar que há pouca relação entre o som e a sílaba impressa na tela do computador, mostrando que quase não há associação entre o estímulo auditivo e a memória visual de sua escrita.

11 No texto de 1976, os autores não mencionam esses estímulos. Eles descrevem a sincronização entre a imagem articulando [ga] e o som [ba]. Essa configuração promove a percepção de [da]. Reproduzo aqui os argumentos utilizados na “troca” de textos.

No segundo, os experimentadores substituíram as sílabas impressas de BA e GA por estímulos manuais (táteis) impressos que simulavam faces produzindo a sílaba /ba/ ou /ga/. Ao contrário do primeiro experimento, neste segundo houve grande efeito das sílabas sentidas sobre as ouvidas. Para eles, então, esses resultados mostram que a abordagem por associação é incompatível com o efeito *McGurk*, mas é tranquilamente acomodada dentro da TRD.

Fowler, em resposta a Ohala, justifica que as evidências fonológicas apresentadas por ele não falseiam a TRD. Quando Ohala diz que se espera que as diferenciações ocorram no sinal acústico, pois a fala assim é veiculada, não há o falseamento do argumento da TDR, de que em realidade os gestos possuem propriedades invariáveis e que elas causam distinções no meio, no sinal acústico. A pesquisadora cita o fato de que, quando há lacunas nos inventários sonoros das línguas, quase sempre se trata de um /p/. Isso justamente ocorre porque /p/ possui *bursts* fracos, i.e., ele não consegue ser “bem” veiculado no ar.

Com referência às dúvidas anedóticas de Ohala, a autora apresenta alguns argumentos interessantes. Para as crianças, o resgate dos gestos do sinal de fala não é apenas biologicamente necessário, é logicamente necessário. Elas precisam recuperar os gestos para aprender a produzir a fala. Esse ponto, embora não apareça claramente na resposta de Fowler, é endossado pelos achados do efeito *McGurk*, que apontam para a preferência na recuperação de aspectos acústicos por crianças, ou seja, aprender a recuperar as articulações seria logicamente necessário para aprender a falar.

Na mesma esteira está o fato de que, por seleção natural, a percepção de eventos sonoros especificados pelo sinal acústico promove a sobrevivência da espécie com mais facilidade do que aconteceria ao ouvir ar estruturado por si só. Para ela, perceber o mundo com base em uma estrutura em mídia é uma função evoluída, mas isso não pode depender de um mapeamento *built-in* de estruturas para eventos, i.e., a fala não seria especial como queriam os adeptos da TM.

Por fim, Fowler pergunta: cornetas e automóveis? Como isso refuta a teoria? Da mesma maneira ela se pergunta (ao considerar o critério *Chicken Little*): se a teoria é extravagante, qual seria a teoria menos extravagante? Para ela, não há!

Diante desse debate, notamos que a escolha de um primitivo para a percepção da fala não é ponto pacífico nos estudos fônicos. A nossa proposta é a de que a natureza dos primitivos de percepção da fala está vinculada à natureza dos primitivos fonológicos, que, por sua vez, são sugeridos pelos cortes epistemológicos propostos pelas teorias fonológicas.

3. FAZENDO A VEZ DE ADVOGADO DO DIABO OU COMO A TM RESPONDERIA À POLÊMICA

Na seção anterior, tratamos das hipóteses e dos argumentos sustentados por Ohala e Fowler com relação à polêmica sobre a natureza ou acústica ou articulatória da percepção da fala. Diante deles, nos perguntamos: quais seriam os motivos pelos quais a polêmica não se desenvolve?

Um dos motivos pode estar vinculado justamente ao fato de cada uma das teorias se fundamentar em seu próprio programa de investigação científica. Ao que parece, esse fechamento tem o objetivo de desenvolver pesquisas que promovam o avanço da teoria. Para tanto, esse olhar estreito às premissas da teoria também promove o desinteresse¹² (epistemologicamente orientado) pelas críticas e ataques dos proponentes de teorias concorrentes, i.e., esse “fechar de olhos” pode fazer com que os proponentes das teorias não se importem com as críticas e discussões levantadas pelos seus adversários, uma vez que tais comentários não seriam interessantes para o programa de investigação aos quais as teorias estão vinculadas.

Diante disso, o desinteresse pode ser definido como a estratégia de não se importar com as críticas e com os desenvolvimentos teóricos e experimentais de seus adversários¹³. É possível verificá-lo na polêmica apresentada na seção anterior, pois Fowler não se “importa” com as críticas de Ohala. Por exemplo, Ohala apresenta evidências sobre a fonologia e as mudanças das línguas para sustentar a proposição de um primitivo de base acústica para a percepção, enquanto Fowler fundamenta sua teoria a partir dos dados sobre o efeito *McGurk*. Nenhum deles apresenta um mesmo fenômeno com outra interpretação. A evasiva predileta nesses casos parece ser a de que os dados apresentados não refutam a teoria, indicando uma falta de interesse sobre o que as críticas do adversário têm a dizer. Para Fowler, seu adversário sequer entendeu qual é o propósito da TRD, a ponto de ela crer que seja necessário “esclarecer” pontos sobre como a TM e a TRD encaram a percepção dos sons da fala de maneira distinta. Sua estratégia é a de rebaixar seu crítico! Seria como dizer que o adversário misturou *alhos com bugalhos*.

12 Não estamos tratando o *desinteresse* como sendo uma “teimosia” dos proponentes de teorias. Estamos considerando o desinteresse como uma estratégia epistemológica para que se consiga desenvolver uma teoria, i.e., interessar-se por toda crítica não promoveria o avanço da teoria, uma vez que seus proponentes a todo momento teriam que revisar suas premissas. Lakatos (1979) tratou dessa questão e a chamou de *princípio de tenacidade*.

13 Lakatos (1979) trataria dessa questão a partir do princípio de tenacidade. Trata-se da tendência de os programas de investigação científica continuarem a se desenvolver mesmo quando se deparam com contraexemplos “recalcitrantes”.

Também é possível identificar um descompasso quanto aos objetivos de pesquisa formulados por cada um dos arcabouços teóricos. A busca pelos traços invariantes das articulações (tal como pretende a pesquisa dentro de um arcabouço dinâmico) não é de interesse de um programa calcado em propriedades acústicas.

A nossa proposta é a de que cada uma das teorias de percepção da fala caminha para uma direção r que, possivelmente, elas não se encontrarão em ponto algum, pois cada fato descoberto por cada um das teorias está em um lugar diferente do mundo. Trata-se de objetos teóricos e observacionais distintos. Seria como se cada uma das teorias tivesse uma lanterna diferente para iluminar as suas descobertas. Por exemplo, uma teoria usa uma lanterna que clareia apenas coisas verdes e outra, coisas laranja. Por si só, essas lanternas trariam à tona objetos distintos. No entanto, podemos notar que cada uma dessas lanternas está apontada para um ponto diferente da realidade. Uma delas é apontada para a água e a outra, para o céu¹⁴. Obviamente, cada uma dessas teorias encontrará coisas distintas em lugares distintos.

É esta relação entre as bases epistemológicas nas quais cada uma das teorias são assentadas, além da falta de interesse pela crítica dos adversários, que caracteriza a dificuldade de se desenvolver a polêmica relacionada aos primitivos de análise.

Por exemplo, é por esse motivo que, mesmo a fonologia estruturalista dos *Preliminaries* não tendo sido levada a cabo, e tendo sido substituída por abordagens gerativistas de fonologia e de percepção da fala (como na TM), teóricos como Ohala ainda buscam dados para verificar quais são os traços linguisticamente relevantes do sinal acústico que promovem distinções nas fonologias das línguas. Isto é, não se abandona a possibilidade de encontrar argumentos relevantes para sustentar o fato de que os primitivos perceptuais são de base acústica.

Ao que parece, o abandono de uma teoria ocorre apenas quando há a diminuição das pesquisas dentro de seu programa de investigação – que, por consequência, acaba perdendo terreno e seguidores. Este parece ser o caso da TM, porque nesta polêmica entre Ohala e Fowler faltou o posicionamento de Liberman, já que a própria Fowler assevera que muito pouco há de comum entre a TM e a TRD.

Para exemplificar que cada uma dessas teorias assenta suas propostas em terrenos epistemológicos distintos, decidimos fazer a vez de “advogado do diabo” e mostrar que a TM apresentava objetivos distintos da TRD e de uma abordagem que assume primitivos de base acústica.

14 Esse ponto é uma adaptação da piada do bêbado apresentada em Borges Neto (2004).

Primeiramente, ao contrário da proposta de base auditiva de Ohala e de base articulatória de Fowler, a TM propõe que a percepção da fala envolve a recuperação dos gestos pretendidos pelo falante. Segundo, a TM assume uma abordagem simbólica sobre a percepção da fala, ao contrário da proposta realista e direta da TRD. Não se trata das articulações que ocorrem, que são implementadas. Para a TM, as articulações recuperadas são os comandos pretendidos pelo falante.

Em outras palavras, a TM parte da assunção de que a percepção recupera os gestos pretendidos e por isso não encontramos uma relação de um para um entre as características acústicas e as articulações realizadas no momento de produção da fala.

Para sustentar sua proposta, a TM apresenta evidências experimentais de que não há relação de um para um entre o sinal acústico e a articulação dos sons da fala. As evidências apresentadas são os achados da década de 1950 com o auxílio do *Pattern Playback* (LIBERMAN et al., 1952, 1955), nos quais não se encontravam características acústicas exclusivas (transição de F2) para o ponto de articulação de consoantes. Ao que tudo indica, as transições assumem valores distintos a depender da vogal do núcleo silábico.

Outra evidência apresentada para proteger as propostas da TM é o fato de a percepção seguir a articulação quando há um desencontro entre a articulação e os sons. Essa assunção é ilustrada pelo efeito *McGurk*. Para a TM, esse efeito é uma prova cabal da existência de módulos e de que, dado o desencontro entre as articulações e os sons promovidos pelo ambiente experimental, a percepção se guia pela articulação. Ou seja, mesmo depois de ter conhecimento sobre o efeito e sobre o modo como o experimento foi conduzido, os sujeitos não conseguem “desfazer” a ilusão. Para a TM, trata-se de uma operação mandatória do módulo fonético que executa ações que recuperam os estímulos distais sem se importar com o que acreditamos ou sabemos.

No que concerne ao que seria interessante de se pesquisar dentro da TM, podemos notar que os estudos desenvolvidos estavam organizados com relação ao seu vínculo com a modularidade da mente (FODOR, 1983), pois serão sugeridas pesquisas que endossem a especialidade da fala em seus vários níveis. Isto é, para o desenvolvimento da proposta era preciso pesquisar os aspectos que faziam da linguagem algo especial. Tanto que os proponentes da TM (LIBERMAN; MATTINGLY, 1985) argumentam que a sua proposta sobre a percepção é totalmente compatível com os achados sobre a sintaxe, por exemplo. Com relação à percepção, os autores propõem o desenvolvimento de mais pesquisas sobre a percepção duplex, as quais poderiam dizer mais sobre a forma como o módulo fonético funciona.

Como podemos ver, a proposta da TM é diferente das de base auditiva e de base articulatória da TRD. Tanto a proposição de primitivos

de análise perceptual quanto as evidências experimentais que sustentam o modelo são distintas. Em outras palavras, enquanto a abordagem auditiva busca por invariantes no sinal acústico e a TRD tenta determinar quais são os invariantes articulatórios que são transmitidos por uma mídia informacional (o ar para a fonação e a luz refletida para as línguas de sinais, por exemplo), a TM busca determinar quais são os aspectos que fazem da fala um “instrumento” especial. Ou seja, para a TM os argumentos de Ohala não invalidariam a proposta motora, uma vez que as evidências fonológicas e de mudança das línguas do mundo poderiam comprovar que as alterações acústicas são detectadas pelo módulo fonético como estímulos proximais e que a percepção (e em seguida a mudança linguística) rastrearía os estímulos distais (a articulação). É preciso lembrar que tais articulações são pretendidas, de modo que os ouvintes perceberiam quais eram as intenções do falante, i.e., recuperar-se-ia a forma subjacente, abstrata.

Ainda mais, os argumentos de Ohala com relação aos sons de motores e à recuperação das articulações que promovem as alterações de pressão do aparelho fonador de um trompetista seriam facilmente respondidos pela afirmação de que não se recuperam as articulações envolvidas nesses contextos pois não se trata da fala. Sendo a fala de caráter especial, o módulo fonético não se daria ao trabalho de “tentar” recuperar as articulações de eventos acústicos que não são reconhecidos como fala.

Ao imaginarmos qual seria o posicionamento da TM diante da polêmica acerca dos primitivos de percepção da fala, não há como não se perguntar por que a TM não entrou na discussão. Ao nosso ver, no momento da polêmica, a TM já poderia ou ter perdido seguidores ou estar ocupada com outra agenda de pesquisa. Tal agenda poderia até mesmo estar relacionada a alguma revisão de pressupostos teóricos, devido ao vínculo a outro programa de pesquisa científica, como o de base dinâmica ou ecológica.

De qualquer forma, é preciso considerar também que pode ter havido o abandono da teoria que decorre da desistência de uma das partes (difícil ocorrer por vontade própria) ou da morte de um de seus proponentes. Este parece ser o caso da TM. O pronunciamento hipotético da TM feito aqui revela que a teoria já não apresentava explicações “convincentes” ou mesmo seguidores. Esse segundo aspecto aproxima as teorias às doutrinas. Esta questão merece atenção, mas não trataremos dela aqui¹⁵.

15 Um agradecimento ao professor doutor Kleber Candioto (PUC-PR) por nos abrir os olhos a esta questão.

4. CONSIDERAÇÕES FINAIS: POR QUE A POLÊMICA NÃO FECHA?

Nas seções anteriores, apresentamos a polêmica em que se envolveram Ohala (1996) e Fowler (1996) sobre a natureza dos primitivos da percepção da fala. Nosso objetivo era “ressuscitar” essa discussão para que tenhamos consciência de que ainda há muita pesquisa a ser feita antes de adotar um primitivo de análise perceptual ou outro.

Também tentamos mostrar os motivos pelos quais a polêmica não fecha. O fato de ela não fechar pode ser considerado um dado importante para a hipótese defendida neste trabalho, a saber: a escolha do primitivo da percepção é dada epistemologicamente, i.e., a depender da teoria fonológica escolhida, é possível contemplar novos fenômenos que, por sua vez, abrem campo para a verificação experimental (a partir de um Programa de Investigação Científica) com relação à produção e à percepção da fala.

Em outro trabalho (NISHIDA, 2012), é possível verificar que cada teoria fonológica propõe novos primitivos de análise que acabam influenciando (ou guiando) as pesquisas sobre a percepção da fala. Tais estudos acabam por corroborar ou refutar conceitos sugeridos pelas teorias fonológicas. Um exemplo da influência dos pressupostos teóricos sobre a pesquisa em percepção da fala é a tentativa de os pesquisadores dos Laboratórios *Haskins* procurarem os invariantes acústicos que guiavam a percepção com o auxílio do *Pattern Playback*. Nesse mesmo exemplo, há o fracasso dessas tentativas enviesadas pelos princípios da Fonologia Estruturalista dos *Preliminaries* (JAKOBSON; FANT; HALLE, 1952).

Diante desse fracasso de se encontrarem os invariantes perceptuais no domínio acústico da fala, Liberman (1957) postula, ainda timidamente, as bases da Teoria Motora da Percepção da Fala (a *Early Motor Theory*). Sua proposta só ganha uma versão mais forte e robusta em 1985, após os avanços propiciados pela Fonologia Gerativa Padrão (CHOMSKY; HALLE, 1968), que possibilita um olhar mais abstrato e mentalista para a produção e percepção da fala. Da mesma forma, só foi possível propor a versão revisada da TM (LIBERMAN; MATTINGLY, 1985) quando dos achados referentes ao efeito *McGurk* (MCGURK; MACDONALD, 1976) e da publicação da teoria modular da mente (FODOR, 1983).

Na “contramão” dos trabalhos de base gerativo-transformacional, há os trabalhos de Fowler (1981 e seguintes) sobre a influência de aspectos coarticulatórios na percepção da fala. É essa abordagem de base dinâmica que promove a proposição da TRD (FOWLER, 1986). Essa teoria passa a trabalhar dentro de um arcabouço teórico de natureza distinta das anteriores. Abandona-se o olhar puramente simbólico (ou dicotomizado pela determinação do que é fonético ou fonológico em uma dada língua), para que os

aspectos reais (de natureza majoritariamente articulatória) envolvidos na produção e percepção da fala sejam considerados. É importante destacar a relevância da proposição do gesto articulatório como primitivo de análise fonológica (BROWMAN; GOLDSTEIN, 1986 e seguintes), pois é com essa abordagem teórica que a TRD dialoga e busca fenômenos passíveis de serem tratados em seu arcabouço teórico.

Diante disso, salientamos, mais uma vez, que cada uma das teorias de percepção da fala mencionada aqui está vinculada a uma teoria fonológica que acaba influenciando a escolha do primitivo perceptual. Se pensarmos estritamente na polêmica mencionada há pouco, podemos notar que Fowler trabalha dentro de uma perspectiva dinâmica de produção e percepção da fala (considerando a produção como um sistema matemático complexo e a percepção como multimodal), enquanto a TM trabalha em uma perspectiva estritamente simbólica. Por sua vez, Ohala (1996) apenas critica as abordagens articulatórias sem se vincular explicitamente a um viés teórico. Em sua argumentação, é possível notar que as evidências apresentadas sugerem que há certa falta de embasamento experimental para tratar dos dados da percepção. Essa falha de natureza empírica que promoveria a implausibilidade (em seus termos) de adotar primitivos de natureza articulatória para a percepção. Talvez seja esse o ponto no qual os adeptos de uma perspectiva dinâmica sobre a linguagem estejam trabalhando. Faltam, contudo, diretrizes que orientem a coleta de dados que embasam empiricamente essa proposta.

Ao resgatarmos trabalhos anteriores de Ohala (e.g., OHALA, 1990), podemos conjecturar que o seu posicionamento contra os primitivos de base articulatória está relacionado ao fato de que seria necessário haver a tradução do nível acústico em um articulatório durante o processo de perceber a fala. Levando em conta, então, que para Ohala (1990) os níveis fonéticos e fonológicos são integrados, não haveria a necessidade de traduzir um nível em outro. Portanto, a proposição de um primitivo de natureza articulatória seria desnecessária. A crítica de Ohala é claramente direcionada à abordagem gerativo-transformacional da TM, uma vez que a da TRD não necessita de tradução de um nível em outro (cf. FOWLER; ROSENBLUM, 1991). A TRD trata o fenômeno perceptual de maneira muito mais ampla. Não há apenas uma relação entre o acústico e o articulatório. De maneira geral, há uma relação entre o articulatório e as mídias de transmissão das características motoras, sejam elas dadas pelo ar, pela luz ou através do tato. Em suma, podemos dizer que se trata de uma abordagem multimodal e nada especial sobre a percepção. Tais questões sequer seriam de interesse para as outras abordagens sobre a percepção da fala.

REFERÊNCIAS

- ALBANO, E. C. *O gesto e suas bordas*. Campinas: Mercado de Letras/FAPESP, 2001.
- BORGES NETO, J. *Ensaio de Filosofia da Linguística*. São Paulo: Parábola, 2004.
- BROWMAN, C.; GOLDSTEIN, L. *Articulatory gestures as phonological units*. *Phonology*, v. 6, p.201-251, 1989.
- _____. Tiers in articulatory phonology, with some implications for casual speech. In: KINGSTON, T.; BECKMAN, M.E. (Eds.). *Papers in Laboratory Phonology I: between grammar and physics of speech*. Cambridge University Press, 1990. p. 341-376.
- _____. *Towards an articulatory phonology*. *Phonology Yearbook*, v.3, p.219-252, 1986.
- _____. *Articulatory Phonology: an overview*. *Phonetica*, v. 49, p. 155-180, 1992.
- CHOMSKY, N.; HALLE, M. *The Sound Pattern of English*. New York: Harper & Row, 1968.
- COLE, R. A.; SCOTT, B. *Toward a theory of speech perception*. *Psychological Review*, n. 81, p.348-374, 1974.
- FODOR, J. A. *Modularity of Mind: An Essay on Faculty Psychology*. Cambridge, Mass.: MIT Press, 1983.
- FOWLER, C. A. *Segmentation of coarticulated speech in perception*. *Perception & Psychophysics*, v. 36, p. 359-368, 1984.
- _____. *An event approach to the study of speech perception*. *Journal of Phonetics*, n. 14, p. 3-28, 1986.
- _____. *Listeners do hear sounds, not tongues*. *The Journal of the Acoustical Society of America*, v. 99, n.3, p. 1730-1741, 1996.
- FOWLER, C.; DEKLE, D. J. *Listening with eye and hand: crossmodal contributions to speech perception*. *J. Exp. Psychol. Hum. Percept. Performance*, v. 17, p. 816-828, 1991.
- FOWLER, C. A.; ROSENBLUM, L. D. Perception of the phonetic gesture. In: MATTINGLY, I. G.; STUDDERT-KENNEDY, M. (Eds.). *Modularity and the motor theory*. Hillsdale: Lawrence Erlbaum, 1991. p. 33-59.
- GIBSON, J. J. *The ecological approach to visual perception*. Boston: Houghton Mifflin, 1979.
- _____. *The senses considered as perceptual systems*. Boston: Houghton Mifflin, 1966.
- JAKOBSON, R.; FANT, G.; HALLE, M. *Preliminaries to speech analysis*. Cambridge: MIT Press, 1952.
- LAKATOS, I. O falseamento e a metodologia dos programas de pesquisa científica. In: LAKATOS, I.; MUSGRAVE, A. (Eds.). *A crítica e o desenvolvimento do conhecimento*. São Paulo: Cultrix, 1979.
- LIBERMAN, A. M. (1957). Some results of research on speech perception. In: LIBERMAN, A. *Speech: a special code*. Cambridge, MA: MIT Press, 1996.
- LIBERMAN, A. M.; DELATTRE, P. C.; COOPER, F. S. *Acoustic loci and transitional cues for consonants*. *Journal of the Acoustical Society of America*, n. 27, p. 769-773, 1955.
- _____. *The role of selected stimulus-variables in the perception of the unvoiced stop consonants*. *American Journal of Psychology*, p. 65, p. 497-516, 1952.
- LIBERMAN, A. M.; MATTINGLY, I. G. *The Motor Theory of Speech Perception Revised*. *Cognition*, v. 21, p.1-36, 1985.

MASSARO, D. *Speech perception by the eye and ear: a paradigm for psychological inquiry*. Earlbaum: Hillsdale, 1987.

MCGURK, H.; MACDONALD, J. *Hearing Lips and Seeing Voices*. Nature, p. 264746-48, 1976.

NISHIDA, G. *A percepção da fala no Estruturalismo*. ReVEL, vol. 8, n. 14, 2010.

_____. *Sobre teorias de percepção da fala*. Curitiba: UFPR, 2012. Tese de doutorado. Inédito.

OHALA, J. *Speech perception is hearing sounds, not tongues*. Journal of the Acoustical Society of America, n. 99, p. 1718-1725, 1996.

_____. *There is no interface between phonology and phonetics*. Journal of Phonetics, n. 18, p.153-171, 1990.

PISONI, D. B. *Auditory and phonetic memory codes in the discrimination of consonants and vowels*. Perception and Psychophysics, n. 13, p.253-260, 1973.

PISONI, D. B.; TASH, J. *Reaction times to comparisons within and across phonetic categories*. Perception & Psychophysics, n. 15, p. 285-290, 1974.

SILVA, A. H. P. *As fronteiras entre Fonética e Fonologia e a alofonia dos róticos iniciais em PB: dados de dois informantes do sul do país*. Campinas: Unicamp, 2002. Tese de Doutorado. Inédito.

Submetido em: 03/09/2014

Aceito em: 15/09/2014